DOCUMENT RESUME

ED 193 288 TM 800 603

AUTHOR van Houten, Carel A.
TITLE Setting Standards of Performance: Towards a Teacher Referenced Test System.
PUB DATE Jun 80
NOTE 24p.; Paper presented at the Fourth International Symposium on Educational Testing (Antwerp, Belgium, June 24-27, 1980). Some print may not reproduce clearly.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Diagnostic Tests; Elementary Education; *Elementary School Mathematics; Foreign Countries; Formative Evaluation; Item Banks; Mathematics Curriculum; Mathematics Materials; *Remedial Mathematics; Summative Evaluation; *Test Construction; Testing Programs; Test Validity
IDENTIFIERS *Netherlands; Teacher Referenced Tests

ABSTRACT

A "teacher referenced" test system is discussed in terms of its development, evaluation of implementation, and final validation. This diagnostic test system for problems in primary mathematics in Holland provides reliable, applicable information about both the learning process for the classroom teacher, and diagnostic instruments and materials used in remediation. One of the early large scale studies involved 29,000 primary school children: findings based on a 10 percent sample revealed that 75 percent of the problems could be diagnosed as a lack of understanding of mathematical concepts: only 25 percent were due to a lack of training. Steps in the test development process included an extensive qualitative orientation phase, and a total of five research and development revision cycles, each including evaluation of the actual school situation and the curriculum content being taught. The test system puts the teacher in charge of setting standards by "pin-pointing" white spots in the curriculum and remediation of pupil problems. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED193288

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SETTING STANDARDS OF PERFORMANCE

Towards a Teacher Referenced Test System

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
C. VAN Houten
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Carel A. van Houten, Ph.D.

Research & Development Department
Wolters-Noordhoff publ., inc.
Groningen
The Netherlands

Paper presented at the Fourth International symposium on Educational Testing, Antwerp, Belgium, june 24-27, 1980.

TM 800603

SETTING STANDARDS OF PERFORMANCE:

Towards a Teacher Referenced Test System

Carel A. van Houten, Ph.D.
Research & Development Department
Wolters-Noordhoff publ., inc.
Groningen, the Netherlands

1. Introduction

1.1 *Setting standards*

In this presentation on the general topic of setting standards of performance, we would like to present a perspective, slightly different from the usual.
Assessing learner performance by testing, and setting standards for this performance has always been based on some kind of compromise. In most cases the standard or cut-off score is set by some quasi-objective artificial statistical technique, or the schooldistrict or government sets some arbitrary standards not based on the performance, but on the percentage of students it can handle in subsequent educational institutions.
The schools (and parents) are becoming more concerned with setting standards of performance as realistic and as close to the learning process and actual school curriculum, as possible.
Therefore our first consideration was to develop a testinstrument for use in actual classroom situations by the usual teacher, which is based on a horizontal assesment of actual performance. This system should provide the diagnostic instruments to put the teacher in charge of setting standards.

1.2 *Using tests in education*

Our second consideration concerns the actual use of tests in the classroom.
Especially during the last decade, the use of norm referenced versus domain referenced test instruments for selection and decision making about the learner and the subsequent learning process, has been widely discussed.
A recent presidential debate by James Popham and Robert Ebel at the AERA 1979 convention seems to prove that both the criterion referenced and the norm referenced tests have their specific uses. However these specific uses are not always recognized by the teachers and educators. Misuse and abuse of test-materials by teachers often occurs, due to lack of training in observing and diagnosing children and to the lack of remediation possibilities.
In this presentation we will discuss some additional possiblilities which might provide a reliable validation standard for a test application on a horizontal basis, which includes the teacher the educational setting and learner specifics as part of development process.

During a 1980 AERA symposium on Evaluation, there were some observations made by the speakers, which illustrate the need for additional test development procedure. Dan Stufflebeam: "The goal of Evaluation in education is not to prove, but to improve. People tend to teach towards (instead on the basis of) certain test results no matter what: with norm referenced as well as with domain referenced tests".
Michael Scriven: (on education) "One should apply analytic evaluation to improve the instruction, by giving the teacher the feed-back and the tools, and the learner the remediation on the basis of the test results".
W. Jim Popham stated that "You as testdeveloper can no longer ignore the instructional needs".

In the following presentation we would like to share with you some of the research we have carried out on this topic.

1.3 *What is new about the 'teacher referenced' test sytem*

Since 1968 we have been involved in developing a diagnostic test system for use by the usual teacher in the usual primary school classrooms.
In the process we included procedures for itemconstruction using the classical item and test construction criteria and criterion referenced methods.

Most importantly, we evaluated the real educational setting and teacher learner decisions, to which the test information should be directed, in the formative stage of the testdevelopment procedure. In addition the actual implemtation of the system in the classroom was analysed. The main objective for the development of this test system was derived from two observations:

1. Misuse and abuse of teachers of existing tests in the schools, because lack of proper training and remediation on the basis of tests.
2. Our own experience in diagnosing and providing materials for individual problems.

In a large scale study with this test in the Dutch Antilles, which was later confirmed in Holland, it was found that more than 70 % of the most common pupils mistakes were based on a lack of understanding the concepts, rather than on a lack of practice and training. Thus in our experience the most common remediation procedure, that is repeating the same problems again and again until mastery, probably is an ineffective way of dealing with learning problems.
These problems and the manner they were delt with provided the main objective for a 'teacher referenced' test system: this system should not only provide reliable and applicable information about the learning process for the teacher, as well as for the learner, but also the diagnostic instruments and materials to remediate on the basis of this information.

1.4 *Characteristics*

Our procedure of testdevelopment contained the following major characteristics:

1. High emphasis on the validity of testitems and remedial materials.
2. Formative Evaluation of teacher and school environment and summative Evaluation of the actual implementation of test information to learning decisions as an integral part of the testdevelopment methodology.
3. In its final form the test system not only provides the relevant test information, but also tools for constructive feed-back and remediation to the teacher and the learner.
4. Evaluation of the implementation of the testsystem. The teacher should be able to apply the information properly.

Given the central role of the teacher in the educational decision making processes, we tentiatively christened this type of test: The "teacher referenced" test system.

2. Overview of the test development process

In the following part we will discuss the steps taken since 1968 to develop a diagnostic test system on primary school arithmetic* skills in cluding remedial materials.
We will show how we aimed to incorporate the school and teaching environment as part of the test development process.

* Arithmetic is equivalent to primary mathematics in Holland.

Inventarisation and first validation

a. *Schoolcurriculum*

On the basis of our own experience in the diagnosis and referal proces of children in special education, a comparison of curriculum objectives content taught at schools, according to age/grade level was derived at for the basic skills at primary school level. By using this set of skills it became possible to evaluate and diagnose a pupil on the basis of his/her personal (test) record and socio economic background, in direct relation with the curriculum content taught. This produced results, which would be consistent for different schools for the same type of pupil. An elaborate prodecure to evaluate the schoolcurriculum content, and learner environment was also included in the later steps of test development.

b. *Educational materials*

In 1968 a first inventarisation of the content and objectives of the three then most widely used educational arithmetic packages was carried out. This inventarisation has been repeated several times during test development and five new educational packages have been added to the validation list.
In addition, in the later steps inventarisations were carried out of the actual use in the classroom of those materials: were parts omitted, or presented in a different manner or in a different order? How much additional training materials were used, etc. These inventarisations were used to analyse the results of the test, as well as to modify and revise existing educational packages !!

c. *Inventarisation of problem areas*

After a few years of diagnosing individual children for counseling and special education a large scale study with some 29.000 primary school – children was carried to pin point the most common problem areas in arithmetic[1].
It was in this study that a first version of a arithmetic test instrument was tried out on a large scale basis. The findings, based on a diagnosis of a 10 % sample, included the observation that, of the problems in primary school arithmetic, 75 % could be diagnosed as lack of basic understanding and only 25 % were found to be based on a lack of training.
Also from this large scale study a sequential/longitudinal line of arithmetic objectives, as they are taught, and subsequently mastered at the different grade level, was constructed. This line was subsequently used to verify the results of the inventarisation of curriculum content and educational materials carried out before.

/4.

STEP 2

Development and formative evaluation

During the 1972-1975 period the objectives and items to be tested were revised from the inventarisation of course materials and the evaluation of curriculum content and context variables. These instruments were further implemented in a large rural/urban school district, in Leeuwarden, the Netherlands.
In addition to this pool of common objectives and test items, remedial teaching 'worksheets' were designed, to match the most common learning problems as diagnosed by the testbattery.
This testbattery in its early form included several commercially available tests, besides the arithmetic test items (item pool).
A diagnosis of total schoolresults in this schooldistrict on the basis of the longitudinal itempool is illustrated in table 2.
The fat line shows how ideally the testresults should be distributed through the 18 testmoments (three tests for each grade) in primary school.
The actual testresults are plotted as small lines in the same table.

STEP 3

Construction of itempool and remedial system

During a four year trial period a revision in an interative mode of the pool of common items, and an inventarisation of common problems was carried out. Also a selection of solutions and remedial materials was tried. During this stage feedback was garnered from the school and teachers and an analysis was made of the actual progress of pupils on the basis of the remedial materials. In addition the experience from the schoolcounseler and school psychologist with experimental versions of the test and remediation materials constitued the basis of the actual construction of an itempool and the remedial teaching materials.

STEP 4

Validation of itempool and remediation

Although in most cases a schoolpsychologist had to be involved to diagnose and execute the remedial program, the information from the teacher, the school and home situations were carefully assessed as controlfactors to explain individual testresults. The individual learner was assessed in a perpective of the actual performance in arithmetic at the school level i.e. information on how does the teacher rate the child; what are other testresults; case history and background. The experimental materials were tried out several years in different schoolsettings. Also an analysis was made to match the testscores on different objectives to the difference in use of course materials. It was only by combining these different sources of information that a valid 'standard of performance' could be derived at for the individual pupil.

/5.

Some majors questions to be adressed were: 1. which testscore signifies an individual problem and should therefore lead to remediation of the pupil 2. At what time a testscore is low for a school or schooldistrict, and should be remediated by training the teacher or changing the curriculum accordingly.

As a final outcome a pool consisting of 1000 arithmetic problems, representing the common objectives from grade 1 to 6 was derived at. Also a remedial teaching program with didactic help for the teacher, as well as additional (training) material for the pupil was constructed. These programs were based on the problems diagnosed by the final testbattery. The process of the administration and diagnosis of the test and the implementation of remediation was still to a great extent a matter of specialists: psychologist and school counselers.
The application of this system provided reliable information on the problems in schools in the district, as well as a diagnosis on individual pupils (for remedial teaching or for special education.)

STEP 5

Revision and construction of the test/itempool

Field tryout

During the course of a two year curriculum development project involving the evaluation of education in 120 primary school classrooms a reliable testinstrument had to be developed.
Given the size of the evaluation study the instruments should yield data which could be processed by computer.
For evaluating the learning results and problems within certain educational settings and using certain course materials, an arithmetic test of 1000 open ended questions was designed, based on the former test instrument.
The items covered the course materials from grade 1 through 6. Therefore a pupil of grade 1 only finished some 30 items where as a sixth grader finished all 1000 of them.
The teacher of the sixth grade would then score some 30.000 open items on an O.M.R. computersheet. We provided a C.M.I. and scoringservice. Each pupil received on the basis of the testresults a detailed computerreport on the particular objectives and test items. In addition a remedial teaching handbook, including most programs from the previous school counseling service, was offered to remediate the weak spots.
This evaluation study provided us with the following information[3]:

- Detailed information on the deficiencies of the course materials in use.
  This information was subsequently applied to develop new course materials.
- Thousands of responses to open items. These were subsequently analysed and used to construct the multiple choice alternatives.
- An evaluation of the need for qualitative test information by the teacher. This has lead to the development of a differentation model for use with course materials. Also the decision to develop a course-independent set of test instruments, for use by the teacher, was based on this evaluation.
  This set can be applied towards a detailled diagnosis of the educational process.
- The inclusion of remedial teaching materials coupled with a diagnostic test battery was found to be a great success.
  A follow up study showed that the actual implementation in the classroom was often lacking: mainly due to the effort by the teacher in first scoring the test, then diagnosing the testresults and finally providing adequate remediation for an average of 30 pupils in the classroom.
- Finally the information through interviews and a follow up questionnaire lead to the notion that the remedial teaching programs as such were found to be valuable. The application in the classrooms was often inhibited by a lack of training in diagnosis and remediation of the teacher, and by organisational problems.
- Diagnosing the different groups on abilities and deficiencies provided information on the concurrent validity of the test.

This information resulted in the objectives for the next evaluation.
- to develop a better scoring model for the test
- te develop an easy-to-read computer diagnosis
- to develop a teacher-friendly remediation system, including hints on classroom organization.

STEP 6

Validation and product evaluation

In 1978, on the basis of the qualitative information from the construction fase a large scale (n = 200 grades/40 schools) tryout of the revised test system was carried out, which included the following objectives[4].

On test development:

- To validate the test itempool and the objectives and match the content to the actual age/gradelevel. This should result in a good concurrent validity and a shorter test per grade.
- To get statistics on the agregate results of the test and the individual items.
- To evaluate if a multiple choice item format should offer a fifth 'open' alternative. The test item could then closer conform to the actual learning process.
- To evaluate two different systems of test scoring by the pupil, for use by young (first grade) children.

On the remedial teaching system:

- To evaluate two different itemformats and computerscoring procedures.
- To study the effects of reporting itemized pupilresults, including systematic mistakes, to the teacher.
- To include information on the organization of remedial teaching in the classroom.

The outcomes of this final formative study resulted in final version of the Rekenmaat test package in conjunction with the E.V.E.R.T. remedial teaching system.
The selection of ± 264 items to be included in the final version was made on the basis of the items statistics in previous experiment.

Itemselection

The final items were selected using the following criteria:

- content validity: items should be similar to curriculum materials used at the different grades
- face validity: teacher should feel confident that the test measures the right objectives
- concurrent validity: items should adequately diagnose deficiencies and abilities for a particular objective
- P value: items should have a reasonable P value (5-10% < p > 95%) and equivalent alternatives.

3. Evaluation of implementation and final validation

So far we discussed the developmental process of the testsystem. In addition the implementation and actual value of the system was evaluated.

STEP 7

Evaluation of implementation and quantitative test validation

After the final revision it was decided that a last check should be carried out to obtain qualitative data on the actual value to the teacher and the implementation of the testsystem in the classroom. In addition, statistical data on the test instrument were collected to validate the test in its final version.

The following elements were evaluated in the final edition:

a. *Final itemformat*
   - multiple choice - alternatives constructed on the basis of diagnosis of most frequent mistakes
   - a fifth 'open' alternative was provided, which sometimes represented the correct answer
   - the 264 items were divided in objectives and then contained in four testbooklets of ± 70 items each.

b. *The revised test scoring system :*

   Including the so called rooftile overlap system tested out before which makes scoring by first and second graders possible.

c. *A revised remedial teaching system containing:*
   1. An elaborate explanation of each individual mistake, remediation programs and hints on classroom organization.
   2. A scoringsystem which provides individual diagnostic information, on the 'total' and on the 'objective' mastered.

Sampling procedure

A group of 90 grades (2800 pupils total N = ± 350 per grade) was selected from all Dutch schools on the basis of five criteria:

1. Rural/urban area.
2. Large school / small school
3. Lower soc.ec. area/higer soc.ec. area
4. Uses major arithmetic course material/uses other materials.
5. Geografic distribution.

All selected school participated, but sometimes teachers showed motivational problems during the two test periods (November and May) in the experiment, and had to be convinced to hang on.

Fieldtest procedure:

The steps in the final fieldtest included:

1. Selecting experimental schools.
2. convincing the teachers to participate (by phone or mail).
3. Visit each schoolteam personally (deliver the testmaterial and explain the test to ensure proper test procedures and expectations.)
4. Conduct a questionnaire type inventarisation of teacher and school situation, to control for variations in curriculum.
5. Keeping in regular contact during the first testperiod.
6. Personally pick up the testforms at the schools, to get a quick evaluation and to ensure proper processing.
7. Reporting of aggregate statistics to school and to developer.
8. 6 months later. Second delivery of test materials to the schools.
9. Telephone contact with school during test.
10. Conduct a questionnaire type survey assessing the teachers opinion on the test and on the format of the test.
11. Processing the testforms (send by mail).
12. Report back individual pupil results including remedial teaching materials to each teacher.
13. Visit each schoolteam to evaluate, and to ensure proper implementation of the testsystem.
14. (To be carried out). Follow up study to find out how well the test system is actual used after 6 months.

Outcome of the final fieldtest based on a 'known group' validation

1. *Content validity*

Results from the inventarisation questionnaire showed that the content of the schoolcurriculum taught is generally consistent with the content of the specific testbook. The aggregate testscores as shown in table 4 illustrate this point. It can also be seen that the test was 'difficult' at the beginning of a new grade, whereas mastery occurred in subsequent testmoments.

2. *Consistency with other tests*

A spearman rankorder correlation coefficient[5] was computed for each school on five indicators. These correlations were accumulated and divided by N.

A comparison was made between the individual testscores on the 'Rekenmaat' test and four other indicators.

/9

- *Testscore - 'teacher grade'.*
  The rankorder coefficient of the total individual testscore 'Bovenbouw' compared with the evaluation of the teacher is an average of .79. Range: highest correlation is .93, lowest .30.
- *Testscore - other (CITO) test.*
  The rankorder coefficient of the total individual testscore 'Bovenbouw' compared with an official schooltest battery is an average of .74. Range: individual variations of .24 upto .92
- *Testscore - readinglevel.*
  The rankorder coefficient of the total individual testscore 'Bovenbouw' compared with the grades on reading show an average of .54 with a range of .25 - .90.
- *Testscore - help needed.*
  The rankorder coefficient of total individual testscore 'Bovenbouw' compared with the observations of the teacher on the amount of help each pupil needed shows a .66 correlation, ranging from .46 to .80. The size of these correlations indicate that the 'Rekenmaat' testsystem provides a valid and adquate instrument for the teacher.

c. *The testing procedure*

As has been reported by my collegue on test in the session on 'anxiety', it was found that the rooftile/overlap system of testbooks, which makes a direct scoring possible, could be applied with young children.
Also a fifth 'open' alternative in addition to the four usual multiple choice alternatives provided for a lower test 'noise' level and a better diagnostic instrument.

d. *The computer analysis and remedial teaching system*

If the reception by these schools is in any way indicative for the general teacher, it seems that the EVERT system is indeed an effective diagnostic and remedial aid to the teacher.

The individual diagnosis coupled with the remedial teaching package was very much appreciated.

e.. *Summary statistics*

The summary statistics of the four tests of 'Rekenmaat' of the fieldtest in november 1979 and in may 1980 are shown in table 3.
In table 4 we have illustrated how the testscores of the same test are distributed through four subsequent test moments.
This table clearly illustrates now the testcontent 'fits' over the four testmoments. Only the teacher knows what has been taught at a particuler moment. Therefore only the teacher can decide on the standard of his/her class given a certain testresult.
To be able to do this, a test should provide detailed information on its content, and the way the group and the individual performs. This was the main reason to develop the EVERT Diagnostic test and remedial teaching system.

/10.

STEP 8

Implementation

The Rekenmaattests and EVERT are now ready for distribution.
The 'Rekenmaat' tests in its final version consists of four different books, each covering upto nine different major objectives in arithmetic/math, primary school level. An overview of the items and objectives is presented in table 5.
The testbook 'Onderbouw' is intented for grade equivalent 1 en 2.
The testbook 'Middenbouw A' and'Middenbouw B' are intended for grade equivalent 3 and 4.
Finally the testbook 'Bovenbouw' is intended for grade equivalent 5 and 6.
In addition to the four testbooks, a testmanual, scoringkeys and scoring-sheets are provided

In addition the specific mistakes and the underlying deficiences are shown. This system can be worked out by hand (with scoring keys), as well as by computer. A latent image version is planned, through which the pupil receives feedback immediately by using chemical treated scoringforms. See also table 6.

For the teacher a special remedial teaching handbook has been made available

In the final table we have illustrated what the 'Rekenmaat' testprocess and the diagnostic system looks like. See table 6.
A follow up implemention study on the actual use of the system will be carried out in november 1980.

4. Summary and conclusions

We first noted some of the reasons of developing a test system with a different perspective. Secondly we have discussed the steps which were necessary to carry out such a test development process.

- These include a very extensive qualitative orientation fase, and a total of five R & D and revision cycles, each including controlled fieldtests. It is also important to note that evaluation of the actual school situation and the curriculum content taught has been very much part of the developmental process.
- The itempool and remedial materials which were developed during eight years in an official setting, were subsequently used for the development of 'Rekenmaat' and the Evaluation and Remedial Teaching system.
- During these last five developmental years a form of 'illumative evaluation', i.e. an evaluation that includes evaluation of the school-setting, the teacher and the learner as an integral part of process, was carried out parallel to the actual test development.
- To provide adequate feedback and low testanxiety with younger children an applied research project using different testsformats and item response modes was carried out. Also experiments were conducted with a number of different computer reporting and processing techniques.
- Finally we discussed the evaluation of the implementation of such a diagnostic testsystem.
  Besides the validity and reliability of the tests and remediation, an important objective for the test development was also the proper application of the test and its results/implications to the usual classroom situation by the usual teacher: Giving the teacher the tools to set his/her own standards and provide remediation materials.

CONCLUSION

In conclusion, we may formulate the following observations on the teacher referenced testsystem:

ADVANTAGES

1. By conducting extensive validation studies at the start of the test development process in close touch with the actual schoolsituation a very good vantage point for the appropiateness of the testsystem is reached.
2. The close contacts between test developer and course material designer in connection with teachers provided a fruitfull base for development.
3. By including the schoolteacher as responsible partner in the test-development a more responsive to real education problems' standard of performance is developed.
4. The testsystem in its final form is no longer a main instrument in selection, but leads to pin pointing white spots in the curriculum and remediation of the pupil problems. The system puts the teachers in charge of setting the standards.
   The teacher can set his/her own standard on the different objectives for the individual pupil and decide if the individual should get he .
6. The testsystem has lead to spinn off development in other areas: such as course materials: C.M.I. systems; other systems.

DISADVANTAGES

1. A long developmental period:
   The test system has taken 12 years to develope.
2. The continuity during such a long and intensive process might constitute a major problem.
   The sponsors, group of authors and projectleaders involved might change several times during this period.
3. It takes a lot of time and effort to motivate teachers in the field tryouts.
4. A large investment in time and money, resulting in a test for just one subject area.
5. The teacher might not be aware of difference in tests.
6. The quality and appropiateness of a diagnostic system like this might not be appreciated. It leads to more work for the teacher.

In final conclusion: we clearly see the advantages of developing and applying a 'teacher referenced' test system but only time can show if the effort involved will pay off.

REFERENCES

1. E. de Groot: 'Rekenen op Curaçao', report to the ministry of Education, 1971.

2. 'Jaarverslag', Schooladviesdienst Leeuwarden, 1973, 1974, 19,75.

3. C.A. van Houten: 'Evaluatie en het ontwikkelen van leermiddelen', Project 615, report presented for V.O.R. group Utrecht, 1977.
   A summary in:
   C.A. van Houten: 'A pragmatic approach to Evaluation in developing curriculum materials', paper presented at A.E.R.A. annual meeting, Toronto, 1978.

4. C.A. van Houten: 'Evaluatie Onderzoek Wiskunde voor de Basisschool, project 616/617, final report, Groningen, 1978.

5. S. Siegel: 'Nonparametric statistics', Mc Graw Hill, New York, 1956.

TABLE I

Overview of the 'teacher referenced' test development process

1968	Step 1 Inventarisation	Inventarisation of schoolcontext and curriculum	Inventarisation of course materials	Inventarisation of learning problems
		Fieldtest Dutch Antilles		
1971	Step 2 Development	Development of hierarchy of schoolcurriculum content Grade 1 - 6	Development of common objectives in course materials	Development of - of individual diagnostic procedures - of remedial teaching materials
		Fieldtest Leeuwarden		
1972-75	Step 3 Revision and construction	Construction of itempool on common objectives		Construction of diagnosing procedures and remedial teaching programs
1976	Step 4 Revision and validation	Revision and validation of itempool		Revision validation of diagnostic and remedial teaching program
		Fieldtest - 615		
1977	Step 5 Construction and revision	Revision and test construction		Revision and construction of diagnostic system
		Large scale Fieldtest 617		
1979	Step 6 Validation and formative evaluation	Validation of two parallel versions of test format and content		Validation of EVERT system
		Revision + Final Fieldtest		
1980	Step 7 Evaluation of implementation Qualitative test results	'Controlled' implementation - user information on Rekenmaat and EVERT diagnostic system - test and item analysis		
1980	Step 8 Implementation and distribution	Implementation and distribution of test system and the computer-scoringservice		

TABLE 2

Schooldistrict Leeuwarden

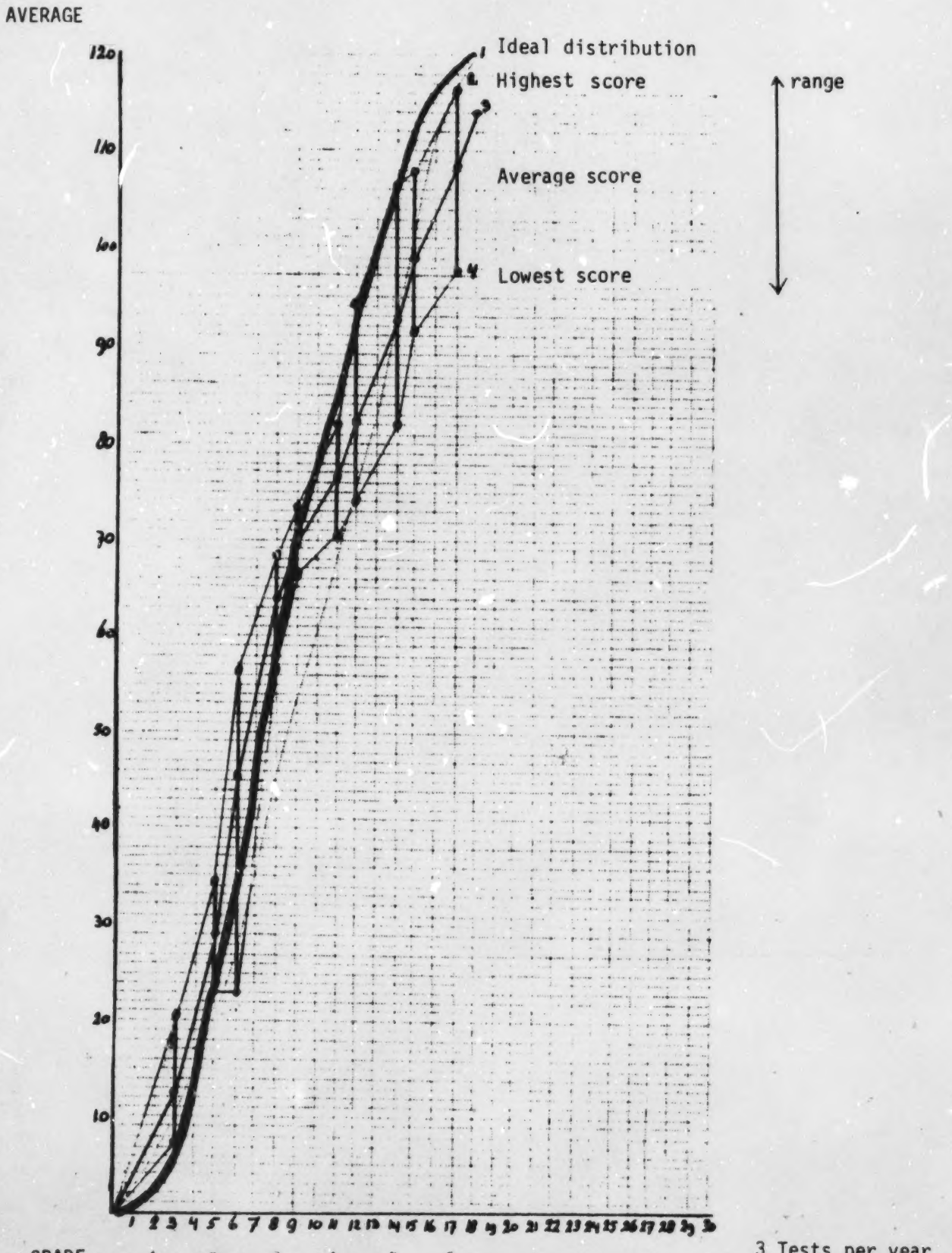
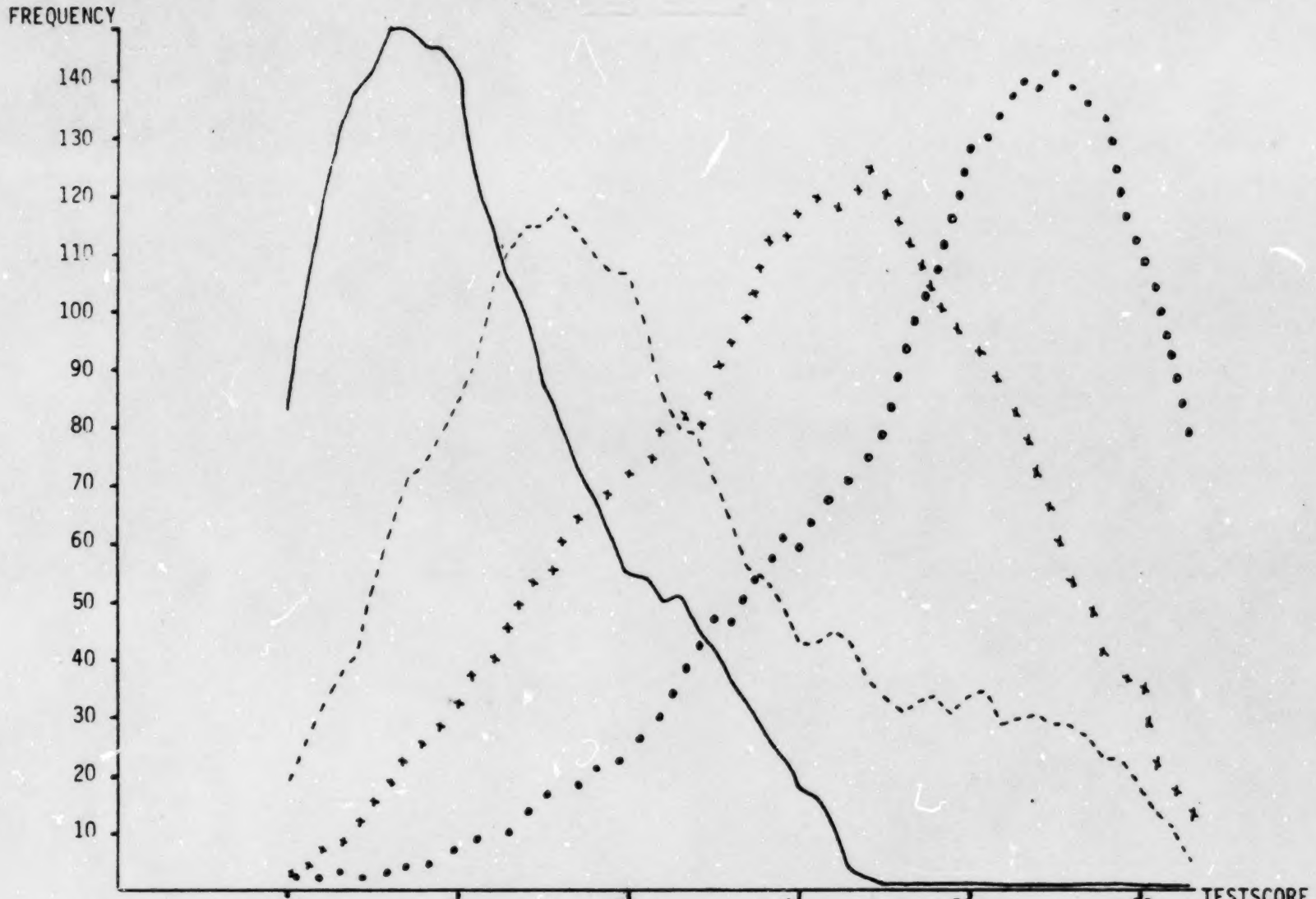


TABLE 3: SUMMARY STATISTICS OF THE FINAL FIELD TEST

Testbook:	ONDERBOUW			MIDDENBOUW A				MIDDENBOUW B				BOVENBOUW			
Grade Date	Grade 1 May '80	Grade 2 Nov.'79	Grade 2 May '80	Grade 3 Nov.'79	Grade 3 May '80	Grade 4 Nov.'79	Grade 4 May '80	Grade 3 Nov.'79	Grade 3 May '80	Grade 4 Nov.'79	Grade 4 May '80	Grade 5 Nov.'79	Grade 5 May '80	Grade 6 Nov.'79	Grade 6 May '80
Sample size	254	354	331	336	315	358	344	331	308	356	340	327	312	319	303
Number of items	70	70	70	68	68	68	68	63	63	63	63	63	63	63	63
Average Score	16.594	31.432	48.795	27.107	40.990	45.095	52.680	15.133	26.536	35.003	45.221	18.667	28.401	37.279	44.452
Standard Deviation	8.559	11.7	13.281	11.093	12.177	10.889	9.299	8.048	12.313	11.359	10.499	6.910	9.491	10.718	11.680
KR 20	.901	.922	.944	.908	.921	.911	.896	.878	.936	.920	.914	.829	.896	.916	.935
Standard Error	2.698	3.2[illegible]3	3.13[illegible]	3.359	3.414	3.257	2.996	2.813	3.121	3.219	3.082	2.855	3.056	3.112	2.976
Number of Subtests	9	9	9	4	4	4	4	5	5	5	5	9	9	9	9

TABLE 4:



Aggregate testscores of four different testmoments

——— grade 3, nov. '79 xxxxx grade 4, nov. '79
- - - grade 3, may '80 ooooo grade 4, may '80

> REKENMAAT, Middenbouw B

TABLE 5:

Items and objectives 'Rekenmaat'

TOETSBOEKJES	REKENDOELEN					
	A	B_1	B_2	B_3	C	D
Onderbouw	*35 items* nrs.: 1 t/m 35	*12 items* nrs.: 36 t/m 47	*2 items* nrs.: 48, 49	*6 items* nrs.: 50 t/m 55	*3 items* nrs.: 56 t/m 58	*4 items* nrs.: 59 t/m 62
Middenbouw A	*32 items* 1 t/m 32	*9 items* 33 t/m 41			*12 items* 42 t/m 53	
Middenbouw B			*24 items* 1 t/m 24	*15 items* 25 t/m 39		*13 items* 40 t/m 52
Bovenbouw			*3 items* 1 t/m 3		*6 items* 4 t/m 9	*6 items* 10 t/m 15
Totaal	*67 items*	*21 items*	*29 items*	*21 items*	*21 items*	*23 items*

						TOTAAL
E	F	G	H	I	J	
2 items nr.: 63, 64				*2 items* nrs.: 65, 66	*4 items* nrs.: 67 t/m 70	*70 items*
					15 items 54 t/m 68	*68 items*
7 items 53 t/m 59				*4 items* 60 t/m 63		*63 items*
13 items 16 t/m 28	*11 items* 29 t/m 39	*9 items* 40 t/m 48	*6 items* 49 t/m 54	*5 items* 55 t/m 59	*4 items* 60 t/m 63	*63 items*
22 items	*11 items*	*9 items*	*6 items*	*11 items*	*23 items*	*264 items*

Objective A :counting, addition, substraction with integral numbers.
Objective B_1 :multiplicationtables.
Objective B_2 :multiplication with integral numbers.
Objective B_3 :deviation with integral numbers.
Objective C :money.
Objective D :measuring.
Objective E :fractions.
Objective F :counting, addition, substraction with decimal fractions (numbers).
Objective G :multiplication and deviation with decimal fractions (numbers).
Objective H :percent.
Objective I :profit and loss.
Objective J :time.

STEPS IN DIAGNOSTIC SYSTEM REKENMAAT.

Step 1 prepare scoring sheet

102496

TEST 0 MA AR 0

SCHOOL KLAS

DATUM dag 10 20 30 1 2 3 4 5 6 7 8 9

mnd jan feb mrt apr mei jun jul aug sep okt nov dec

jaar 79 80 81 82 83 84 85

14 13 12 11 10 9 8 7 6 5 4 3 2 1

SCHOOLNUMMER

KLASNUMMER

LEERLINGNUMMER

BEOORDELING

slecht zwak onbeslist goed groot

Step 2 insert scoring sheet in test book

OM TE OEFENEN:

Welke vrucht

ligt naast de schaal?

1

2

51867

44367

1897348

15097448

.....

4963

3042 x

Step 3 handscoring or computerscoring

Step 4 Diagnosis (see next page)

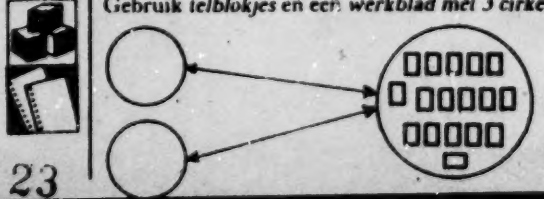
Step 5 remedial teaching

0-10

10 = 17 - . .

Van de hoeveelheid 17 wordt zoveel afgetrokken tot er 10 overblijft. Ook hier gaat het weer om inzicht in getalesplitsen.

Gebruik *telblokjes* en een *werkblad met 3 cirkels verbonden door pijlen.*



De 17 blokjes worden in de grote cirkel gelegd. (zie ook voorafgaande beschrijvingen bij deze methodiek).

EVALUATIE EN REMEDIAL TEACHING BASISONDERWIJS
WOLTERS-NOORDHOFF B.V. GRONINGEN

E.V.E.R.T Diagnostic system

SAMPLE OUTPUT

Computer Analysis for the teacher

K L A S S E - O V E R Z I C H T

VAK	: REKENEN
TOETSBOEKJE	: ONDERBOUW
TOETSNAAM	: REKENNAAT
SCHOOL	: 632
KLAS	: 1
AANT. LEERL.	: 32
AANT. ITEMS	: 70

GEMIDDELDE SCORE PER L.L.	: 11.3
STANDAARD-AFWIJKING	: 7.9
TOTAAL GOEDE ANTWOORDEN	: 343
TOTAAL NIET GEMAAKT	: 143e
TOTAAL FOUTE ANTWOORDEN	: 401

* TABEL 1 *

S Y S T E E M F O U T E N - A N A L Y S E

SYSTEEMFOUT	AANT. LL.	LEERLINGNUMMERS
1	14	4 9 10 11 12 17 19 20 21 22 23 24 26 31
2	21	2 4 5 6 8 9 10 11 12 15 16 20 21 22 23 24 25 30 31 32 34
3	5	9 10 12 22 30

* TABEL 3 *

EVALUATIE EN REMEDIAL TEACHING BASISONDERWIJS

I T E M - S C O R E S , ONDERBOUW

LEERDOEL	ITEM	PERCENTAGES P= EN A=ANTWOORDEN A	B	C	D	E	O	ABSOLUTE P= EN A=ANTWOORDEN A	B	C	D	E	O	LEERLINGNUMMERS MET FOUT/GEEN ANTWOORD
A	1	0	90*	0	0	0	10	0	19*	0	0	0	2	11 20
	2	19	10	0	0	67*	5	4	2	0	0	14*	1	2 5 6 9 13 17 30
	3	0	86*	0	14	0	0	0	18*	0	3	0	0	3 8 17
	4	0	10	86*	5	0	0	0	2	18*	1	0	0	5 12 17
	5	81*	5	0	0	10	5	17*	1	0	0	2	1	12 13 14 19
	6	10	5	0	0	86*	0	2	1	0	0	18*	0	3 6 12
	7	5	10	57*	14	0	14	1	2	12*	3	0	3	2 5 9 14 16 18 19 20 22
	8	14	43*	5	5	24	10	3	9*	1	1	5	2	2 3 4 9 10 12 13 14 15 18 19 23
	9	52*	0	5	5	19	19	11*	0	1	1	4	4	2 3 4 7 9 10 13 16 18 21
	10	5	0	10	82*	5	19	1	0	2	13*	1	4	3 4 6 9 13 14 15 22
	11	5	5	67*	0	19	5	1	1	14*	0	4	1	5 9 10 14 16 19 22
	12	14	19	5	5	14*	43	3	4	1	1	3*	9	2 3 4 5 6 8 9 10 11 14 15 16 17 19 20 27 29 31 32
	13	0	5	5	52*	10	29	0	1	1	11*	2	6	2 3 8 9 10 12 13 14 16 21
	14	14*	0	14	29	14	29	3*	0	3	6	3	6	2 3 4 7 8 9 10 11 12 13 14 18 19 24 26 27
	15	5	0	38*	0	14	43	1	0	8*	0	3	9	2 3 4 5 8 9 10 12 15 16 18 28
	16	5	24*	5	5	5	57	1	5*	1	1	1	12	2 3 4 9 7 8 9 10 12 15 16 17 18 29 30
	17	86*	5	0	10	0	0	18*	1	0	2	0	0	3 16 22
	18	0	33	5	33*	14	14	0	7	1	7*	3	3	2 3 4 6 7 8 10 12 13 14 17 20 21 24
	19	0	0	10	43*	14	33	0	0	2	9*	3	7	2 4 8 9 12 13 15 16 17 20 21 24
	20	29	5	0	10*	10	48	6	1	0	2*	2	10	2 3 4 5 6 7 8 9 10 13 14 15 17 19 20 21
	21	0	14*	14	14	10	48	0	3*	3	3	2	10	3 4 6 7 8 9 10 11 12 13 14 16 17 18 19 20 21
	22	0	5	0	5*	5	86	0	1	0	1*	1	18	2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

>>

* TABEL 4 *

EVALUATIE EN REMEDIAL TEACHING BASISONDERWIJS

O V E R Z I C H T P E R L E E R L I N G , ONDERBOUW

(ZIE HOOFDSTUK VOORGESTELD DOOR DATE STEENBEEK)

SCHOOL : 622 KLAS : 1 VAK : REKENEN , AANTAL ITEMS : 70

BLADNR	LEERLING -	NIVEAU	AANTAL ITEMS	ANALYSE PER LEERDOEL ITEMS FOUT EN/OF NIET GEMAAKT
202541	.17	2	GOED : 25	TOTAAL 36 A 2A 3D 4A 12D 14D 16C 18C 19E 20A 21 23A 26A 31A 32A 33C
	BARBARA T		FOUT : 36 NIET : 9 TOT : 1 2 3	B1 34E 37 39 40D 41D 42 43B 46 66E 67D 34 B2 EN B3 48E 49 50A 51E 52C 53A 54 55 17 C EN I 56D 57D 58A 60D 8 D, E EN J 59B 60C 61A 62A 63C 64A 70A 20 36